**Jing Huang, Jing Liu, Wenbing
Tao, Zhenxing Yang, Rui Qiu,
Shaoning Yu and Chaoneng Ji***State Key Laboratory of Genetic Engineering,
Institute of Genetics, School of Life Sciences,
Fudan University, Shanghai 200433, People's
Republic of China

Correspondence e-mail: chnji@fudan.edu.cn

Received 13 January 2013

Accepted 10 April 2013

Crystallization and preliminary X-ray analysis of the CRP–cAMP–DNA (full length) complex

The *Escherichia coli* cyclic AMP receptor protein (CRP) is a well known transcription activator protein. In this study, CRP was overexpressed, purified and cocrystallized with cAMP and a 38 bp full-length double-stranded DNA fragment. The full-length segment differed from the half-site fragments used in previous crystallization experiments and is more similar to the environment *in vivo*. CRP–cAMP–DNA crystals were obtained and diffracted to 2.9 Å resolution. The crystals belonged to space group $P3_121$, with unit-cell parameters $a = b = 76.03$, $c = 144.00$ Å. The asymmetric unit was found to contain one protein molecule and half a 38 bp full-length double-stranded DNA fragment, with a Matthews coefficient of $2.62 \text{ Å}^3 \text{ Da}^{-1}$ and a solvent content of 53.14%.

1. Introduction

The *Escherichia coli* cyclic AMP receptor protein (CRP), which is also referred to as the catabolite gene activator protein (CAP), is a classical model for structural and mechanistic studies of transcription activation. Upon cAMP binding, CRP interacts with specific DNA sites in or near promoter sequences. After this interaction, CRP recruits RNA polymerase to the DNA to ultimately activate transcription (Busby & Ebright, 1999).

CRP is a 47 kDa homodimer (Aiba *et al.*, 1982; Cossart & Gicquel-Sanzey, 1982). Each subunit consists of 209 amino acids that form two domains. The larger N-terminal domain, which is composed of amino acids 1–133 (Fic *et al.*, 2009), is responsible for subunit dimerization and cAMP binding. The smaller C-terminal domain is composed of amino acids 139–209 and is responsible for DNA binding through a characteristic HTH motif (Brennan & Matthews, 1989). CRP was the first transcription regulatory protein to have its three-dimensional structure determined (McKay & Steitz, 1981). Structures of unligated CRP (Sharma *et al.*, 2009), of CRP complexed with cAMP (McKay *et al.*, 1982; Weber & Steitz, 1987; Passner *et al.*, 2000) and of CRP complexed with cAMP and DNA (Schultz *et al.*, 1991; Parkinson *et al.*, 1996; Passner & Steitz, 1997) have been determined using X-ray crystallography.

All of the DNA sequences in these previous crystal structures of CRP complexes have a consensus binding site (TGTA) for CRP. In addition, the sequences of these DNA strands were not completely identical and included 30 bp (Parkinson *et al.*, 1996; Chen *et al.*, 2001), 31 bp (Schultz *et al.*, 1991), 38 bp (Napoli *et al.*, 2006) and 46 bp (Passner & Steitz, 1997) fragments. However, the double-stranded DNAs in these structures are broken (with a single-phosphate gap) in each strand and include two symmetric fragments, which are referred to as half-site sequences (Schultz *et al.*, 1990). Each of the half-site sequences has a 4 bp self-complementary overhang at the 3' end; these hybridize with each other to provide completely symmetric binding sites in each molecule of the protein dimer. The purification of long sequences is difficult because it requires a great quantity of DNA (approximately 1 mg) for crystallization experiments and the symmetric sequences easily form hairpins through internal self-complementation, which results in the half-site sequences that have universally been used in previous experiments. Because it is widely known that interaction between protein and DNA is far more complex than was originally imagined, a single mutation in an amino

© 2013 International Union of Crystallography
All rights reserved

acid or a deoxyribonucleotide may have a great influence on the structure and function of the biological molecule. Thus, the question needs to be answered: can two symmetrical half-site sequences really replace a full-length fragment? To obtain complex crystals that are closer to the real structure, we carefully selected DNA sequences with a full-length sequence of 38 bp (Napoli *et al.*, 2006) and conducted the following experiment.

2. Materials and methods

2.1. Expression and purification of CRP

CRP was overexpressed and purified as previously reported (Tao *et al.*, 2010). The cells were resuspended in 250 ml buffer A (50 mM Tris-HCl, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 5% glycerol pH 7.8) and disrupted by pressure (JN-3000 Plus, JNBIO, People's Republic of China) at 277 K. The lysate was successively passed through a pre-equilibrated Bio-Rex 70 column (Bio-Rad), a hydroxyapatite column (Fluka) and a Phenyl Sepharose column (GE Healthcare).

We collected the effluent liquid based on the results of SDS-PAGE (Fig. 1) and a Bradford assay and then dialyzed the effluent overnight in TEK100 buffer (50 mM Tris-HCl, 100 mM KCl, 1 mM EDTA, 1 mM DTT pH 7.8). The CRP solution was then injected into a centrifuge filter (Millipore) and centrifuged at 5000g and 277 K for 5 min. The centrifugation step was repeated until the protein concentration was greater than 20 mg ml⁻¹ and the concentrated protein was kept at 203 K until use in crystallization.

2.2. DNA preparation

Through sequence analysis and the use of *BLAST*, the fragments that were synthesized and used for crystallization in previous studies were found to have a single-phosphate gap, as described in §1. We chose two sequences that are commonly used in crystallization. Two types of oligodeoxyribonucleotide were then synthesized (Invitrogen), a 31 bp DNA fragment (5'-GCGAAAAGTGTGACATATG-TCACACTTTTCG-3') and a 38 bp fragment (5'-ATTTTCGAAAA-ATGTGATCTAGATCACATTTTCGAAAT-3'), and dissolved in buffer (5 mM sodium cacodylate, 200 mM NaCl, 5 mM EDTA pH 7.4). The strands were annealed by heating the solution to 363 K followed by cooling to 295 K for 12–15 h. Molecular exclusion was used, including the use of Superdex 75 (Qiagen), to separate double-

stranded hybridized fragments from self-complementary sequences. An experiment was designed to confirm the affinity between CRP and these DNA fragments. The band for the complex with the 38 bp fragment is more obvious than that for the other fragment, indicating that the 38 bp sequence has a stronger affinity than the shorter fragment (Fig. 2). Therefore, a large quantity of the 38 bp full-length oligodeoxyribonucleotide was synthesized and purified by PAGE (Invitrogen). The double-stranded hybridized fragments were ultra-filtered to obtain a final concentration of 2 mM and the DNA was stored at 253 K until use in the crystallization step.

2.3. Crystallization and X-ray data collection

The CRP, cAMP and DNA were first mixed at a molar concentration ratio of 1:2.5:0.75. The CRP concentration was approximately 20 mg ml⁻¹. The complex solution was then placed on ice for 2 h. The preliminary crystallization conditions were screened by sitting-drop vapour diffusion at 293 K using Index HT and Crystal Screen HT (Hampton Research, USA). Crystals were obtained by mixing 1.2 µl complex solution with 1.2 µl reservoir solution and equilibrating the drop against 100 µl reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.2 M ammonium acetate, 45% (v/v) (±)-2-methyl-2,4-pentane-diol.

Prior to data collection, a single crystal was soaked in cryoprotectant solution (25% glycerol) and flash-cooled directly in a liquid-nitrogen stream at 100 K.

The X-ray diffraction data were collected using a MAR DTB detector system at an X-ray wavelength of 0.97876 Å on beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF). The complete data set included 180 images of 1° oscillation. The data were processed with the *HKL*-2000 package using 100 images (Otwinowski & Minor, 1997).

3. Results and discussion

In this study, recombinant CRP protein was successfully over-expressed in *E. coli*. After purification and screening of crystallization conditions, crystals of the recombinant protein were obtained using a

Figure 1

A Coomassie Blue-stained 15% reducing SDS-PAGE gel showing the purity of the CRP used for crystallization. Lane 2, molecular-weight marker (labelled in kDa). Lane 1, CRP sample loaded onto the phenyl Sepharose column. Lanes 3–13, CRP sample purified by the Phenyl Sepharose column; the concentration in lanes 8–10 was greater than 20 mg ml⁻¹.

Figure 2

The agarose gel results showing the affinity between CRP and DNA. Lane 1, DNA marker (labelled in bp); lane 2, 38 bp dsDNA sample (band C); lane 3, mixed sample composed of CRP, cAMP and 38 bp dsDNA (band B); lane 4, mixed sample composed of CRP, cAMP and 31 bp dsDNA (band A); lane 5, 31 bp dsDNA sample (band D).

reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.2 M ammonium acetate, 45%(v/v) ($\pm$)-2-methyl-2,4-pentanediol (Fig. 3). The preliminary crystallographic data statistics for the CRP complex crystal are given in Table 1 along with global indicators of the quality of the diffraction data (Weiss, 2001). The crystal belonged to space group $P3_121$, with unit-cell parameters $a = b = 76.03$, $c = 144.00$ Å. The Matthews coefficient of $2.62 \text{ Å}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 53.14%, indicated that there was one CRP molecule in the asymmetric unit.

To confirm that the crystal contained protein, DNA and cAMP, we soaked a crystal in reservoir solution to eliminate the complex solution outside the crystal and dissolved it in 10 μl H_2O . The results of agarose gel electrophoresis (Fig. 4) indicated that the crystal contained both DNA and protein.

According to previous experiments, CRP does not interact with the specific DNA until cAMP binds (Busby & Ebright, 1999). All of the 15 CRP structures which contain DNA also contain cAMP molecules. In five of the 15 structures each CRP molecule binds to two cAMP molecules and in the others each CRP molecule binds to one cAMP molecule. Therefore, we have reason to believe that the crystal contains protein, DNA and cAMP.

Figure 3
The crystal of CRP-cAMP-DNA (38 bp full-length). The crystal was grown in a reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.2 M ammonium acetate, 45%(v/v) ($\pm$)-2-methyl-2,4-pentanediol. The dimensions of the fusiform crystal are approximately $1.2 \times 0.3 \times 0.3 \text{ mm}$.

Figure 4
The agarose gel results showing the composition of the crystal. Lane 1, DNA marker (labelled in bp); lane 2, crystal sample composed of protein and DNA (band A); lane 3, 38 bp dsDNA sample (band B); lane 4, mixing solution used for crystallization.

Table 1

Statistics of the collected data.

Values in parentheses are for the outermost resolution shell.

Source	Beamline BL17U, SSRF
Wavelength (Å)	0.97876
Resolution range (Å)	50–2.9 (2.95–2.90)
Space group	$P3_121$
Unit-cell parameters (Å)	
$a = b$	76.03
c	144.00
Observed reflections	
Total	65269
Unique	11148
Completeness (%)	99.8 (100.0)
Molecules per asymmetric unit	1
V_M ($\text{Å}^3 \text{ Da}^{-1}$)	2.62
$R_{\text{merge}}^\dagger$ (%)	8.60 (54.6)
Average $I/\sigma(I)$	31.96 (3.35)
$R_{\text{r.i.m.}}^\ddagger$ (%)	9.4 (56.8)
$R_{\text{p.i.m.}}^\S$ (%)	3.9 (22.8)
Mosaicity ($^\circ$)	0.631
Solvent content (%)	53.14

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum over all reflections, $\sum_i$ is the sum over i measurements of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity of all i observations of reflection hkl . $^\ddagger R_{\text{r.i.m.}} = \sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $N(hkl)$ is the redundancy. $^\S R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

As the full-length DNA interacts with two CRP molecules according to the interaction pattern, there may be two or four cAMP molecules in our CRP-DNA (38 bp full length) complex structure. In addition, in the CRP-DNA structure with the same DNA (38 bp half-site) sequence in the PDB, each CRP molecule binds to one cAMP molecule. Therefore, it is reasonable that the CRP-DNA (38 bp full length) complex binds two cAMP molecules. However, this cannot be confirmed until the structure has been determined.

In future studies, we will attempt to determine the structure of the CRP-cAMP-DNA complex using the molecular-replacement method. The crystal structure may provide more precise information on the mechanism of the cAMP-induced allostery that is required for DNA binding and transcription. We predict that there may be some distinct differences between the two complex structures because the DNA fragments used in this experiment were full-length sequences and not half-site sequences as used in previous experiments. However, the work will still be meaningful even if we do not discover any significant differences when the new structure is obtained, as it will support the previously identified structures. In addition, the method of employing half-site sequences can be extended to the assembly of an oligomer of any length to explore crystallization of a protein with a very long DNA sequence.

We thank the staff of beamline BL17U at SSRF for their help during data collection. We also acknowledge Professor Gong Weimin at the Institute of Biophysics of the Chinese Academy of Sciences. This work was supported by the National Basic Research Program of China (2009CB825505), the National Natural Science Foundation of China (30770427), the New Century Excellent Talents in University (NCET-06-0356), the Shanghai Leading Academic Discipline Project (B111) and the National Talent Training Fund in Basic Research of China (J0630643).

References

- Aiba, H., Fujimoto, S. & Ozaki, N. (1982). *Nucleic Acids Res.* **10**, 1345–1361.
- Brennan, R. G. & Matthews, B. W. (1989). *J. Biol. Chem.* **264**, 1903–1906.
- Busby, S. & Ebright, R. H. (1999). *J. Mol. Biol.* **293**, 199–213.
- Chen, S., Gunasekera, A., Zhang, X., Kunkel, T. A., Ebright, R. H. & Berman, H. M. (2001). *J. Mol. Biol.* **314**, 75–82.

- Cossart, P. & Gicquel-Sanzey, B. (1982). *Nucleic Acids Res.* **10**, 1363–1378.
- Fic, E., Bonarek, P., Gorecki, A., Kedracka-Krok, S., Mikolajczak, J., Polit, A., Tworzydło, M., Dziedzicka-Wasylewska, M. & Wasylewski, Z. (2009). *J. Mol. Microbiol. Biotechnol.* **17**, 1–11.
- McKay, D. B. & Steitz, T. A. (1981). *Nature (London)*, **290**, 744–749.
- McKay, D. B., Weber, I. T. & Steitz, T. A. (1982). *J. Biol. Chem.* **257**, 9518–9524.
- Napoli, A. A., Lawson, C. L., Ebright, R. H. & Berman, H. M. (2006). *J. Mol. Biol.* **357**, 173–183.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., Ebright, R. E. & Berman, H. M. (1996). *J. Mol. Biol.* **260**, 395–408.
- Passner, J. M., Schultz, S. C. & Steitz, T. A. (2000). *J. Mol. Biol.* **304**, 847–859.
- Passner, J. M. & Steitz, T. A. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 2843–2847.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1990). *J. Mol. Biol.* **213**, 159–166.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Science*, **253**, 1001–1007.
- Sharma, H., Yu, S., Kong, J., Wang, J. & Steitz, T. A. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 16604–16609.
- Tao, W., Li, F., Liu, H., Bao, X., Gong, W. & Yu, S. (2010). *Acta Cryst.* **F66**, 516–519.
- Weber, I. T. & Steitz, T. A. (1987). *J. Mol. Biol.* **198**, 311–326.
- Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.